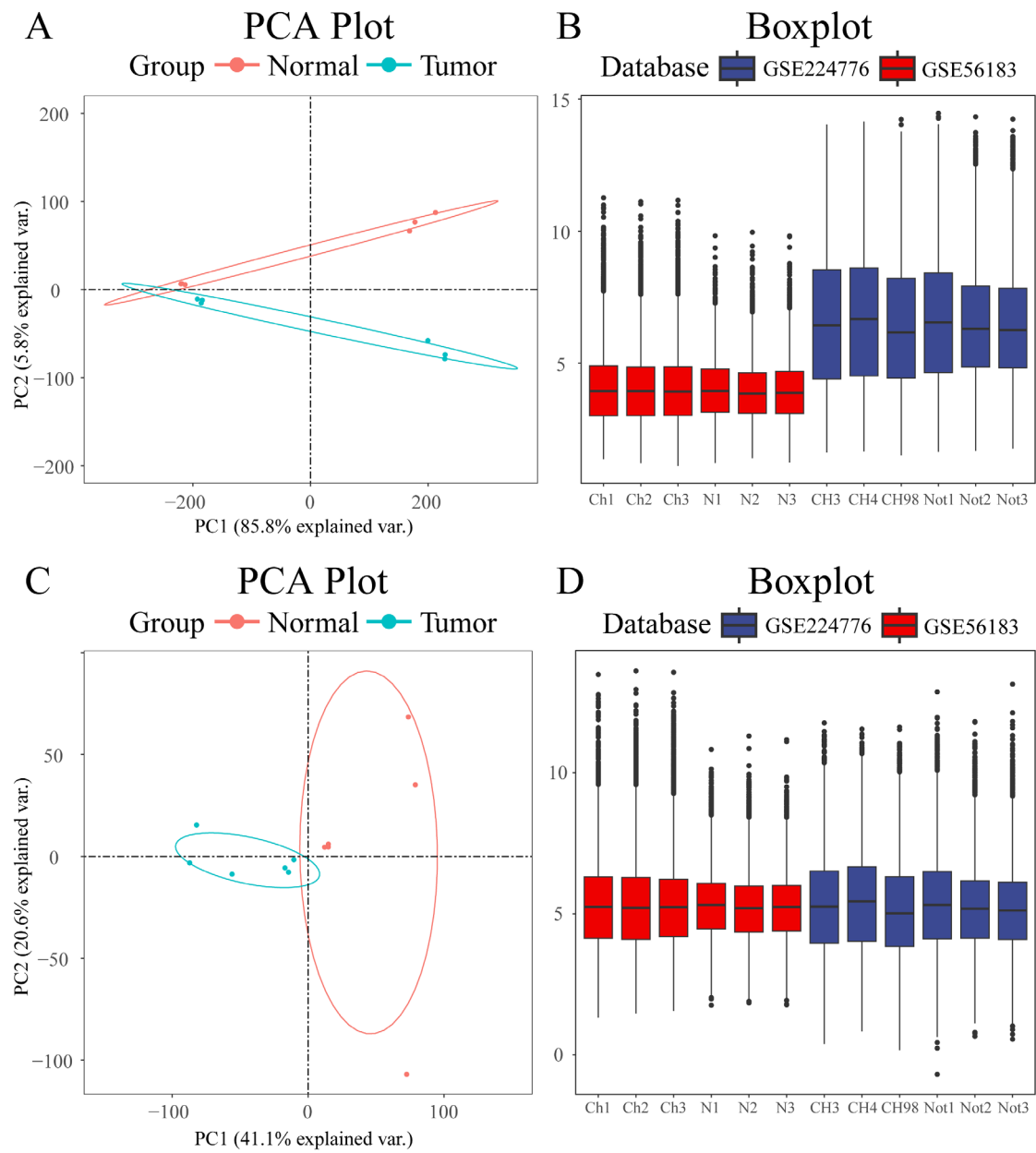
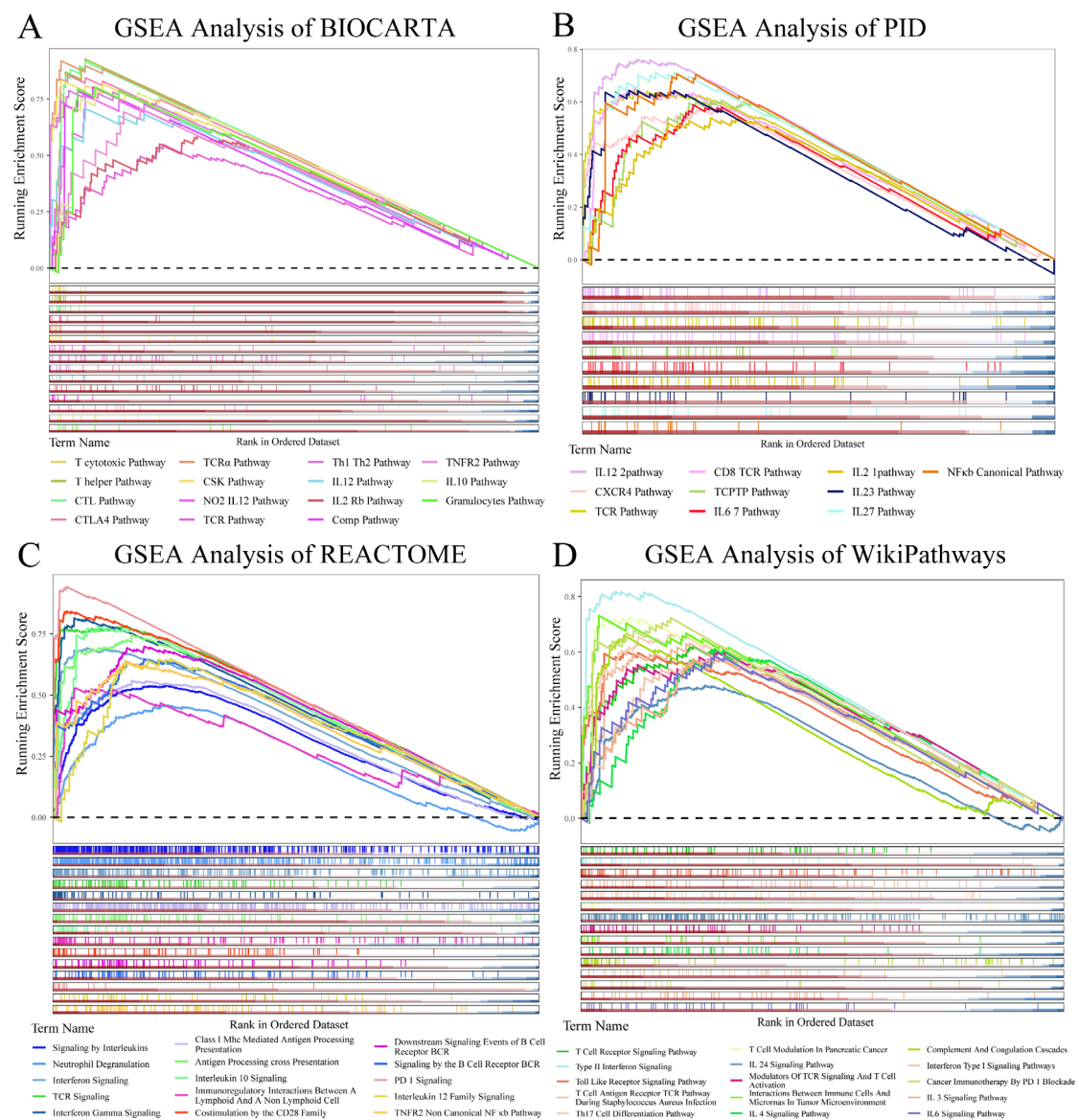


**Figure S1 | Batch effect removal and data integration of microarray datasets.** **(A)** principal Component Analysis (PCA) plot before batch effect removal showing clear separation between normal and tumor samples, with PC1 explaining 85.8% and PC2 explaining 5.8% of variation. **(B)** Box plots showing expression distribution across samples from different datasets (GSE224776 and GSE56183) before batch correction. **(C)** PCA plot after batch effect removal demonstrating successful integration of datasets, with PC1 explaining 41.1% and PC2 explaining 20.6% of variation. **(D)** Box plots showing normalized expression distribution after batch effect removal.

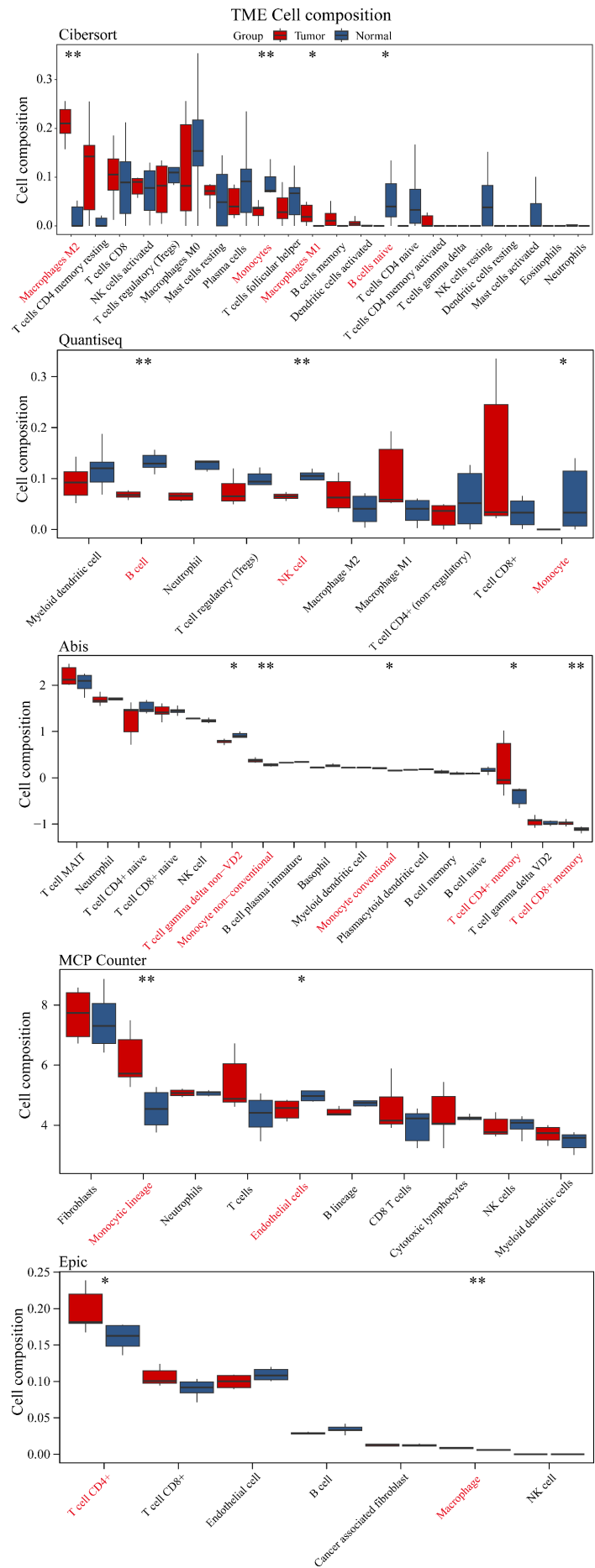


**Figure S2 | Extended pathway analysis reveals comprehensive immune-related signatures. (A)** GSEA analysis using BIOCARTA database showing enrichment of T cell-related pathways including T cytotoxic, T helper, CTL, and TCR signaling pathways. **(B)** GSEA analysis using PID database demonstrating enrichment of IL-12, CD8 TCR, IL-2, and NFκB canonical pathways. **(C)** GSEA analysis using REACTOME database highlighting enrichment of interferon and antigen processing pathways. **(D)** GSEA analysis using WikiPathways showing enrichment of T cell receptor signaling and cytokine-related pathways.

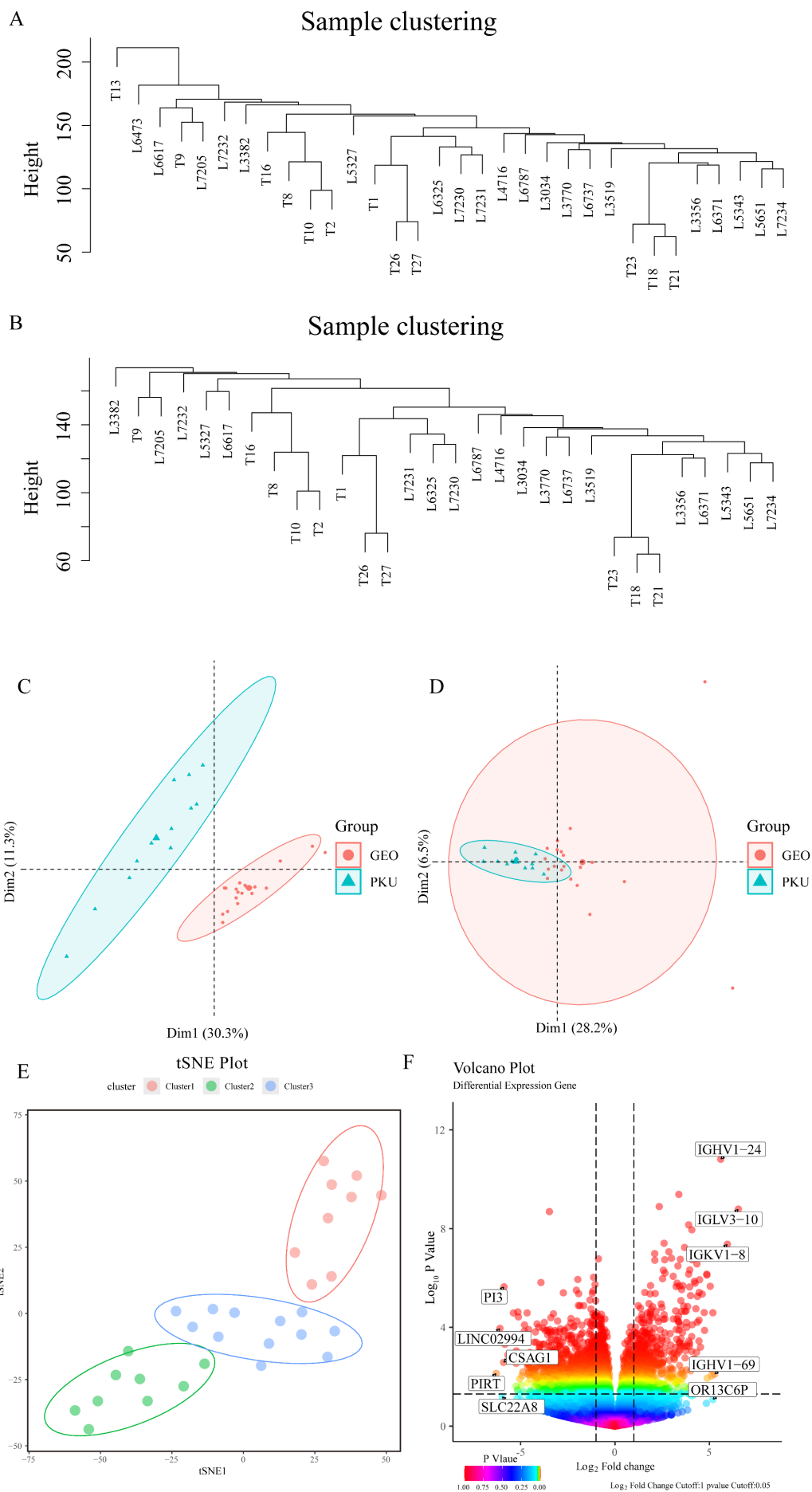


**Figure S3 | Multi-algorithm validation of immune cell infiltration patterns.**

Comprehensive immune cell composition analysis using five different computational approaches. From top to bottom: Cibersort showing distinct patterns of immune cell infiltration between tumor and normal samples; Quantiseq demonstrating differential immune cell populations; Abis revealing significant differences in T cell and macrophage populations; MCP Counter highlighting monocyte lineage differences; and Epic showing distinct T cell and macrophage patterns. Red boxes indicate tumor samples, blue boxes indicate normal samples. Statistical significance:  $*P < 0.05$ ,  $**P < 0.01$ .



**Figure S4 | Sample clustering and differential expression analysis.** (A,B) Hierarchical clustering dendrograms showing sample relationships before (A) and after (B) outlier removal. (C,D) Principal component analysis plots showing sample distribution from different sources (GEO and PKU) before (C) and after (D) batch effect correction. (E) tSNE plot showing three distinct clusters identified through consensus clustering. (F) Volcano plot of differential expression analysis between clusters, with significantly altered genes highlighted ( $|\log_2 \text{fold change}| > 1$ , adjusted  $P < 0.05$ ).



**Figure S5 | Network and enrichment analysis demonstrates interconnected immune**

**pathways in chordoma. (A)** Network visualization of enriched GO Biological Process terms.

Node size represents pathway size, color intensity indicates normalized enrichment score

(NES), and connecting lines show relationships between pathways. The network reveals

major clusters centered around lymphocyte-mediated immunity, T cell activation, and

leukocyte differentiation, with oxidative phosphorylation showing distinct separation. **(B)**

KEGG pathway network analysis displaying interconnected immune signaling pathways. The

TCR pathway forms a central hub connected to IL-12, NFAT TF, BCR, Th1/Th2, and IL-2

STAT5 pathways. Node size corresponds to pathway size, and color intensity indicates NES

values. **(C)** GSEA analysis using BIOCARTA database showing running enrichment scores

for multiple immune-related pathways, including T cytotoxic, T helper, TCR, and IL

pathways. **(D)** GSEA analysis using PID database demonstrating enrichment patterns of

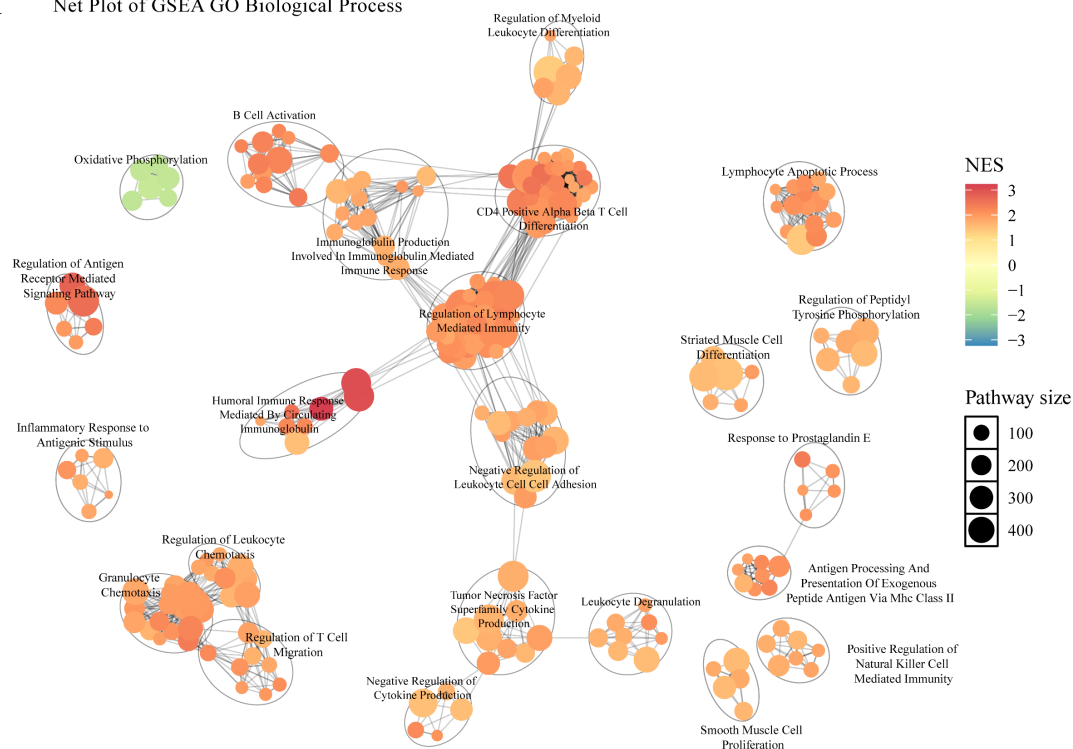
immune signaling pathways, particularly in TCR, IL-12, and NFκB signaling cascades. Both

C and D include gene set position plots (bottom) showing the distribution of pathway

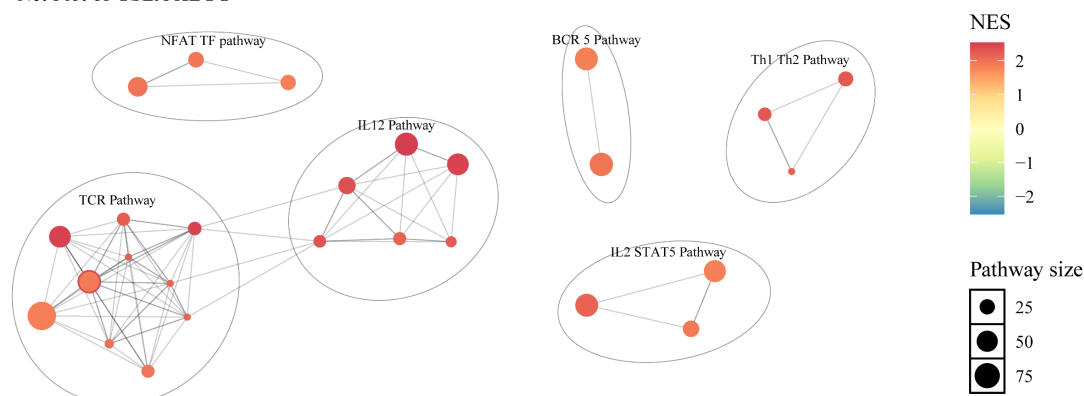
members across the ranked dataset.



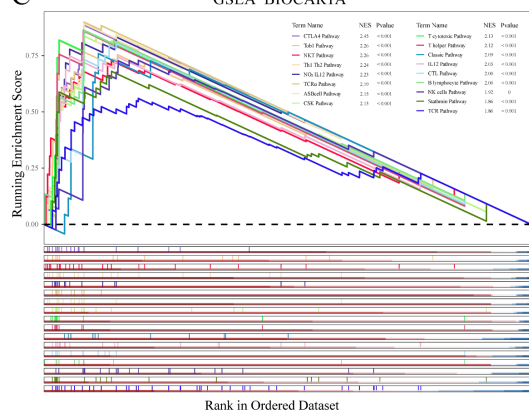
**A** Net Plot of GSEA GO Biological Process



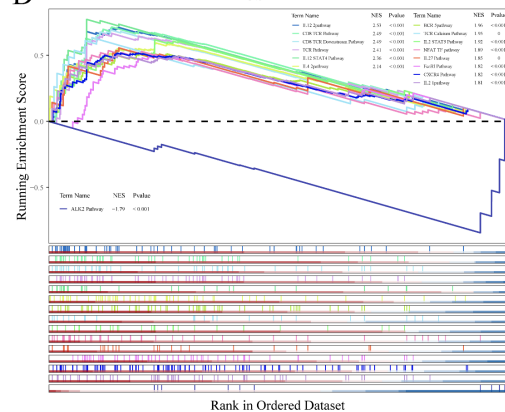
### B Net Plot of GSEA KEGG



C GSEA BIOCARTA



D GSEA PID

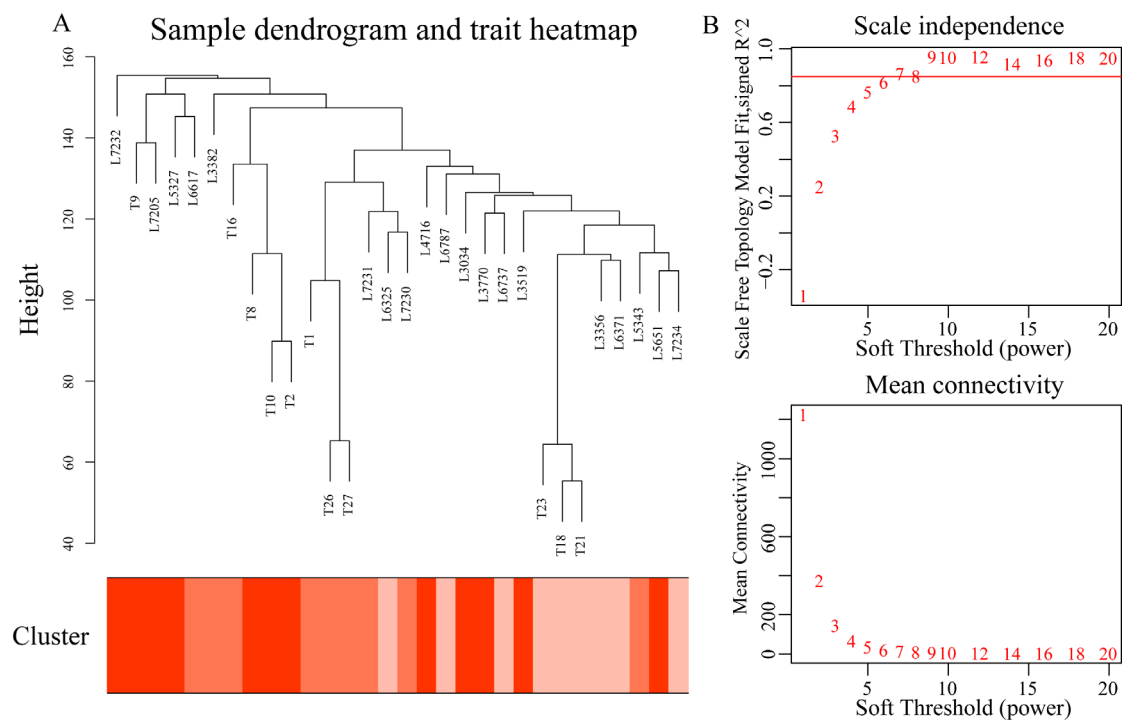


## Figure S6 | Construction of a scale-free co-expression network using WGCNA. (A)

Sample dendrogram displaying hierarchical clustering of chordoma tissues (rows) based on their gene expression patterns. Each vertical branch line corresponds to a single sample, and the grouping patterns reflect global expression similarities. The heatmap below shows the assigned cluster memberships, with varying shades of red indicating distinct subgroups identified by dynamic tree cutting.

(B) Plots depicting the selection of a soft-thresholding power ( $\beta$ ) to achieve an approximate scale-free topology in the gene co-expression network.

The top panel shows the scale-free topology model fit ( $R^2$ ) for different  $\beta$  values; a horizontal red line indicates the selected threshold ( $R^2 = 0.85$ ). The bottom panel displays mean connectivity (degree) as a function of the soft-threshold power. Increasing  $\beta$  reduces overall connectivity, ensuring that the resulting network exhibits scale-free properties. In this analysis,  $\beta=7$  was selected as the optimal power setting.



**Table Legends**

**Table S1 | Comprehensive GSEA Results for Hallmark, C2, and C5 Gene Sets.** Columns include pathway identifiers (ID, Description), gene set size (setSize), enrichment score (enrichmentScore), normalized enrichment score (NES), significance metrics (pvalue, p.adjust, qvalue), rank, leading\_edge genes, and core\_enrichment. These data outline the pathways enriched in chordoma versus normal tissues.

**Table S2 | Immune Cell Infiltration and Immune-Related Analyses.** This multi-sheet file provides outputs from different immune deconvolution algorithms (ABIS, CIBERSORT, EPIC, MCP-counter, quanTIseq, xCell), the ESTIMATE tool, and results for immune checkpoints, immune-related genes, and immunophenocore scoring. Each sheet shows infiltration levels or gene-set scores for chordoma samples.

**Table S3 | Immune Checkpoints and Immune Gene Distributions Across Chordoma Samples.** Rows (e.g., T18, T1, T21, T23, T26, T27, T9) represent individual chordoma samples or subgroups, with columns listing expression or abundance values for various immune checkpoints and immune-related genes, highlighting differences across samples.

**Table S4 | Enrichment Analysis of Cluster-Specific DEGs (C2 and C5 Databases).** Columns include ID, Description, setSize, enrichmentScore, NES, pvalue, p.adjust, qvalue, rank, leading\_edge, core\_enrichment, Database, and status. These metrics reflect the significance of enriched pathways derived from cluster-level differential expression in chordoma.

**Table S5 | GO and KEGG Functional Annotation of Core T Cell-Related Genes.**

Columns (Description, GeneRatio, BgRatio, pvalue, p.adjust, qvalue, geneID, Count) provide key statistics for enriched Gene Ontology (GO) terms and KEGG pathways linked to T cell function. This table highlights the biological processes in which identified hub genes are involved.